زرمهر، فاطمه؛ منصوری، علی؛ کارشناس، حسین (۱۴۰۰). مدل‌سازی موضوعی و کاربرد آن در پژوهش‌ها: مروری بر ادبیات تخصصی.
پژوهش‌نامه کتابداری و اطلاع‌رسانی، ۱۱ (۱)، ۳۹-۲۳.
DOI: 10.22067/infosci.2021.24128.0

مدل‌سازی موضوعی و کاربرد آن در پژوهش‌ها: مروری بر ادبیات تخصصی

فاطمه زرمهر[۱]، علی منصوری[۲]، حسین کارشناس[۳]

تاریخ دریافت: ۹۹/۲/۶ تاریخ پذیرش: ۱۴۰۰/۴/۱۰ نوع مقاله: مروری

چکیده

مقدمه: مدل‌سازی موضوعی یکی از تکنیک‌های متن‌کاوی است که امکان کشف موضوعات نامعلوم در مجموعه اسناد، تفسیر اسناد بر اساس این موضوعات و استفاده از این تفاسیر برای سازماندهی، خلاصه کردن و جستجوی متن‌ها را به‌طور اتوماتیک میسر می‌کند. آشنایی با مفهوم و تکنیک مدل‌سازی موضوعی، و کاربرد آن در کشف موضوعات و سازمان‌دهی منابع اطلاعاتی از اهداف اصلی این پژوهش است.

روش‌شناسی: پژوهش حاضر از نوع کتابخانه‌ای است که در آن، ضمن معرفی مدل‌سازی موضوعی، به دسته‌بندی و مرور کاربردهای این تکنیک بر اساس ماهیت عملکردی آن و ارائه نمونه تحقیقاتی که از این تکنیک استفاده نموده‌اند پرداخته است.

یافته‌ها: الگوریتم‌های مدل‌سازی موضوعی علاوه بر سه هدف اصلی مبنی بر کشف موضوعات پنهان، تفسیر اسناد بر اساس موضوعات و نهایتاً سازمان‌دهی و طبقه‌بندی متون، در کشف موضوعات و روابط پنهان در حوزه‌های علوم، بازیابی اطلاعات، دسته‌بندی مدارک بر اساس موضوعات، کشف الگوهای برجسته و رویدادهای در حال ظهور، خوشه‌بندی مفاهیم حوزه‌های علمی، تحلیل سیر تحول مفهومی در طول دوره‌های تاریخی، تعیین روابط سلسه‌مراتبی مفاهیم یک حوزه یا زمینه خاص علمی و غنی‌سازی فهرست واژگان کاربرد دارد.

نتیجه: مدل‌سازی موضوعی با تکیه بر یادگیری ماشین و بهره‌گیری از دانش هوش مصنوعی به‌عنوان یکی از رویکردهای نوین سازماندهی منابع اطلاعاتی مطرح شده و مطالعات جدی در این زمینه در حال انجام است. لذا با کاربرد الگوریتم‌های مدل‌سازی موضوعی در راستای خودکارسازی استخراج موضوع و کشف موضوعات نهان موجود در منبع می‌توان بر تقویت و روزآمدسازی نظام‌های نوین سازمان‌دهی منابع اطلاعاتی عمل کرد.

کلیدواژه‌ها: متن‌کاوی، مدل‌سازی موضوعی، کشف موضوع، سازماندهی اطلاعات، تشخیص موضوع

۱. دانشجوی دکتری علم اطلاعات و دانش‌شناسی، دانشگاه اصفهان، ایران، fahim2398@gmail.com

۲. دانشیار گروه علم اطلاعات و دانش‌شناسی، دانشگاه اصفهان، ایران، (نویسنده مسئول)، a.mansouri@edu.ui.ac.ir

۳. استادیار گروه مهندسی هوش مصنوعی، دانشگاه اصفهان، ایران،h.karshenas@eng.ui.ac.ir

مقدمه

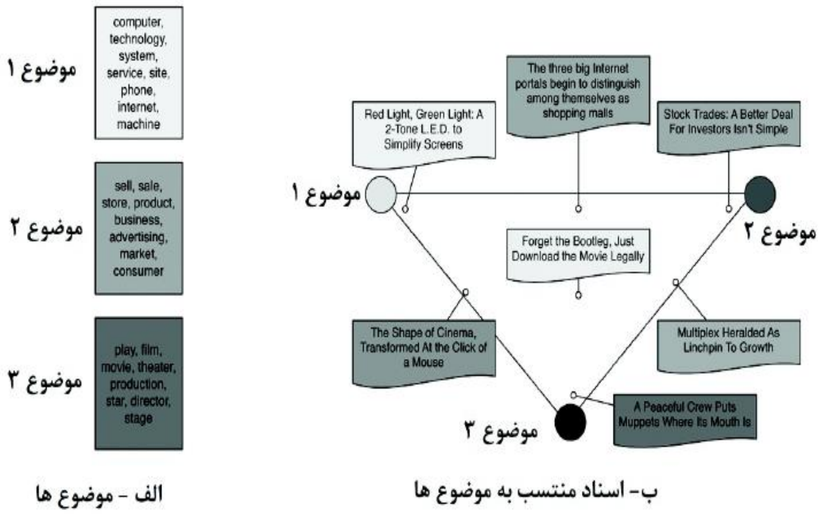
با توجه به ظهور فضای دیجیتال و گسترش رو به رشد تولید متون در این محیط، برای مدیریت بهتر حجم عظیم مدارک الکترونیکی نیاز به استفاده از تکنیک‌ها و رویکردهای تجزیه‌وتحلیل جدیدی است که بتواند به‌صورت خودکار سازمان‌دهی، جستجو، نمایه‌سازی و مرور مجموعه‌های بزرگ را میسر نماید (Hwang,Wei, Lee, & Chen, 2017)؛ به همین دلیل، پژوهش‌های زیادی در این رابطه صورت گرفت (Baeza-Yates & Ribeiro-Neto, 1999; Arora, Kanjilal, & Varshney, 2010; Nadezhda & Aleksey, 2018). روش پایه پیشنهادشده توسط پژوهشگران حوزه بازیابی اطلاعات در مورد مجموعه‌های عظیم مدارک، کاهش حجم مدارک درون مجموعه به برداری از اعداد حقیقی است که هرکدام دال بر بسامد تکرار کلمات بود. برای غلبه بر این مشکل، پژوهشگران بازیابی اطلاعات اخیراً از تکنیک‌های متن‌کاوی همچون الگوریتم‌های مدل‌سازی موضوعی استفاده می‌کنند (Blei & Jordan, 2003).

مدل‌سازی موضوعی یکی از اشکال تجزیه‌وتحلیل متن به‌منظور بررسی رابطه بین کلمات داخل مدرک است، جایی که کلمات در کنار یکدیگر تشکیل‌دهنده درون‌مایه و موضوع اصلی متن است. ایده مدل‌سازی موضوعی هم‌راستا و منطبق با باور بسیاری از زبان شناسان و فیلسوفان است، همان‌طور که ویتگنشتاین[1] (۲۰۱۰) معتقد است که معانی از ارتباط بین کلمات حاصل می‌شود و نه جایگاه مستقل کلمه. به‌عبارت‌دیگر، معنی کلمه از بافت و زمینه حاصل می‌شود. مدل‌سازی موضوعی کلمات مشابه ازنظر معنایی را به هم مرتبط می‌کند. به‌دلیل تأکید مدل‌سازی موضوعی بر همبستگی[2] مشکل چندمعنایی کلمه[3] و یا مترادفات[4] را نیز بر اساس زمینه‌ای که در آن استفاده می‌شود حل می‌نماید. برای نمونه، کلمه «شیر» می‌تواند با کلمات دیگری مانند «مسمومیت»، «لبنیات»، در یک دسته‌بندی موضوعی حضور داشته باشد و یا هم رخدادی آن با کلمات دیگری چون «حیوان»، «وحشی»، «درنده» در دسته‌بندی موضوعی دیگر، معانی متفاوت این کلمه را نشان دهد.

اولین تلاش برای توسعه مدل‌سازی موضوعی توسط پاپادیمیتریو، تاماکی، راگاوان و ومپالا[5] (۱۹۹۸) و هافمن[6] (۱۹۹۹) انجام شد. بعدها این تکنیک با پژوهش‌های بیشتر از سوی بلای و جوردن[7]

1. Wittgenstein
2. Correlation
3. Polysemy
4. Synonomy
5. Papadimitrio, tamaki, Raghavan, & Vempala
6. Hofmann
7. Blei & Jordan

(۲۰۰۳) توسعه یافت. الگوریتم‌های مدل‌سازی موضوعی روش‌های آماری هستند که کلمات داخل یک متن را تحلیل کرده و از این طریق موضوعات داخل متون را استخراج می‌کنند. همچنین ارتباط این موضوعات با یکدیگر و نیز تغییر آنها در طول زمان را مشخص می‌کنند. هدف مدل‌سازی موضوعی تجزیه‌وتحلیل مدارکی است که به موضوعات مختلفی پرداخته است و موضوعات، احتمالی از توزیع کلمات است (Alghamdi & Alfalqi, 2015) (شکل ۱). مدل‌سازی موضوعی رویکردی است که موضوع‌های درون پیکره‌ای از اسناد را استخراج می‌کند و به هر سند برخی از این موضوع‌ها را نسبت می‌دهد. بنابراین یک مدل موضوعی، دو برونداد دارد. نخستین برونداد توزیع موضوع بر روی واژگان است که در قالب دسته‌های موضوعی نشان داده می‌شود. برونداد دیگر یک مدل موضوعی، امکان انتساب یک سند به چند موضوع مرتبط است (مسعودی و راحتی‌قوچانی، ۱۳۹۴، ۱۱۷). شکل (۱) سه موضوع استخراج‌شده از اسناد مجله خبری نیویورک‌تایمز را نشان می‌دهد. هر واژه با درصدی احتمال، در یک موضوع حضور دارد. همچنین این شکل نشان می‌دهد چگونه یک سند می‌تواند با یک یا چند موضوع شناخته شود.



شکل ۱. مثالی از توزیع کلمات بر روی موضوع‌ها و توزیع موضوع‌ها بر روی اسناد در پیکره
(Boyd-Graber, Blei, & Zhu, 2007) **New York Times**

بنابراین، برخلاف روش‌های طبقه‌بندی سنتی، یا روش‌های خوشه‌بندی، که مدرک یا داده را تنها در یک دسته‌بندی قرار می‌دهد، مدل‌سازی موضوعی، امکان تعلق مدرک به طبقات و دسته‌های مختلف را بر اساس درجه عضویت آن میسر می‌نماید.

ماهیت عملکردی مدل‌سازی موضوعی

مدل‌سازی موضوعی اساساً مبتنی بر رابطه ترایایی[1] یا رابطه انتقالی و هم رخدادی[2] واژگان است (Kontostathis & Pottenger, 2006). در ریاضیات، بین سه عضو a و b و c از مجموعه A یک رابطه ترایا برقرار است هرگاه بتوان از وجود رابطه دوتایی بین a و b از یک‌سو، و b و c از سوی دیگر، نتیجه گرفت که a و c نیز دارای همان رابطه هستند (Blei & Lafferty, 2007).

هم رخدادی شیوه‌ای از تحلیل محتوا است که از طریق هم رخدادی واژه‌ها یا مفاهیم در متون و مدارک عمل می‌کند. تحلیل هم رخدادی نمونه‌ای از روش مدل‌سازی گرافیکی است که در آن از اندیشه‌های مربوط به تحلیل رابطه استفاده می‌شود. لذا این روش ابزاری قدرتمند در کشف دانش است (Neff & Corley, 2009; Liu, Hu, & Wang, 2012). علاوه بر تعاریف فوق، تعریف‌های دیگری توسط پژوهشگرانی همچون (Kostoff, Eberhart, Toothman, & Foo, 1997; Ding & Chowdhhury, 2001; Rokaya, 1997) ارائه شده است که به‌صورت مختصر شامل تعاریف زیر است:

۱- نوعی تحلیل محتوا برای کشف الگوها و تعیین گرایش‌های موضوعی
۲- تعیین درجه ارتباط بین مفاهیم و اصطلاحات
۳- کشف شبکه مفاهیم حوزه‌های علمی
۴- کشف روابط مفهومی میان مدارک
۵- تحلیل سیر تحول مفهومی در طول دوره‌های زمانی و ... .

از طرفی روش‌های مختلفی برای دسته‌بندی و خوشه‌بندی و تشخیص موضوعات منابع علمی و غیرعلمی استفاده می‌شود که هر کدام اشکالات خاص خود را دارد. یکی از اشکالات، عدم‌تشخیص موضوعات پنهان و میزان تخصیص هر موضوع به هر مدرک است. روش‌های متن‌کاوی از جمله مدل‌سازی موضوعی به اذعان پژوهش‌های انجام شده یکی از روش‌های قوی برای حل این مسئله است. با توجه به اینکه تخصیص موضوع برای کتابداران یکی از دغدغه‌های اصلی است، انتخاب روشی مناسب ضرورت دارد، لذا در پژوهش حاضر با استفاده از روش کتابخانه‌ای اقدام به بررسی بنیادی این روش و معرفی ویژگی‌ها، قابلیت‌ها و کارکردهای آن نموده است.

1. Transitive Relation
2. Co-occurrence

مرور نوشتارها در تبیین کارکردهای مدل‌سازی موضوعی

اگرچه تکنیک مدل‌سازی موضوعی روشی آماری است و ابتدا در حوزه علوم کامپیوتر شکل‌گرفته است اما کاربرد آن در اکثر شاخه‌های علوم ازجمله علوم طبیعی، پایه و انسانی در حال اثبات است. با توجه به سه هدف اصلی تکنیک مدل‌سازی موضوعی مبنی بر کشف موضوعات پنهان، تفسیر اسناد بر اساس موضوعات و نهایتاً سازمان‌دهی و طبقه‌بندی متون می‌توان نوشتارهای مربوط به این حوزه را نیز به سه دسته عمده تقسیم نمود. بااین‌حال استفاده از این تکنیک فراتر از سه هدف مذکور در حال پژوهش و ارزیابی است. از این‌رو، در ادامه ضمن مطالعه و مرور پژوهش‌های انجام‌شده با سه هدف یادشده به سایر کاربردهای این تکنیک با استناد به پژوهش‌های انجام‌شده خواهیم پرداخت.

۱ـ کشف موضوعات و روابط پنهان در حوزه‌های علوم: از دیدگاه اسپارک جونز[1] (۱۹۷۱) دو کلمه که در یک متن یا یک گفتمان مشترک به‌کاررفته شود می‌توانند به هم مرتبط باشند. به‌عبارت‌دیگر، آنها به‌دلیل حضور در منبعی یکسان ممکن است ازنظر معنایی به هم مرتبط باشند. همسو با دیدگاه ویکری و ویکری[2] (۱۹۸۷) نیز اگر به مجموعه تمام جملات موجود در یک متن بیندیشیم می‌توانیم دریابیم که چه ساختار ارتباطی پیچیده‌ای دارد. درون جمله، واژه‌ها در طبقات نحوی جای می‌گیرند. در ورای این، جملات نیز به شیوه‌ای با یکدیگر ارتباط می‌یابند که بتوانند بحث موجود در یک متن را پیش برند. درنتیجه با مجموعه‌ای از مفاهیم آشکار و پنهان در درون‌متن مواجه هستیم. هاگن[3] (۲۰۱۸) در مطالعه‌ای به تجزیه‌وتحلیل محتوا و موضوعات اصلی دادخواست‌های الکترونیکی به کمک مدل‌سازی موضوعی پرداخت. هدف وی ارتقا و ارزیابی مدل‌های LDA بود. نتایج پژوهش‌های او نشان داد که موضوعات تولیدشده در مدل‌سازی موضوعی نسبت به موضوعات برگرفته از تجزیه‌وتحلیل دستی این مزیت را دارد که زمینه‌های مختلف مطرح در متن را بیان نموده که ممکن است از دید تحلیلگران موضوعی پنهان مانده باشد. از دیگر پژوهش‌هایی که به بررسی استفاده از تکنیک مدل‌سازی موضوعی در کشف موضوعات پنهان و نامعلوم در متون پرداخته و در ضمن تحول موضوعات را در طول زمان نشان داده است را می‌توان در آثار (Newman, Hagedorn, Chemudugunta, & Smyth, 2007; Steyvers, Smyth, Rosen-Zvi, Griffiths, 2007; Cain, 2016; Figuerola, García & Pinto, 2017; Fang, Yang, Gao & Li, 2018; Hagen, 2018; Sanandres, Madariaga, Abello, 2018; Bitterman, & Fischer, 2018؛ محمدیان،

1. Sparck Jones
2. Vickery & Vickery
3. Hagen

۱۳۹۳) مشاهده کرد. هدف‌های ضمنی که از پرتو این دسته از پژوهش‌های حاصل‌شده، کشف موضوعات یا مسائل اصلی و به‌روز در یک حوزه مشخص، تشخیص مسائل کمتر پرداخته‌شده ولی نیازمند بررسی بیشتر، کشف موضوعات اصلی پژوهش‌ها در یک بازه زمانی مشخص، و یا تغییر روند پرداختن به یک موضوع یا موضوعات در طول زمان و کشف موضوعات موردعلاقه خوانندگان هست.

۲ـ بازیابی اطلاعات: همواره برای مدیریت بهتر حجم عظیم و رو به رشد مدارک الکترونیکی نیاز به استفاده از تکنیک‌ها و رویکردهای تجزیه‌وتحلیل بوده که بتواند به‌صورت بهینه به بازیابی مؤثر اطلاعات منطبق با نیاز کاربر منجر شود. در همین راستا تکنیک مدل‌سازی موضوعی به‌عنوان روشی کارآمد در کاهش بعد مدارک و کنترل حجم عظیم مدارک، بازنمون جنبه‌های مختلف موضوعی در یک مدرک یا پایگاه و بیان نسبت موضوعات پرداخته‌شده در هر مدرک معرفی شد است. همچنین امکان بررسی میزان ارتباط بین مدارک با توجه به شباهت موضوعی آنها فراهم شد که به‌نوعی به بازیابی مؤثر اطلاعات مرتبط بود. کین[1] (۲۰۱۶) در مطالعه‌ای امکان استفاده از مدل‌سازی موضوعی در ارتقا دسترسی به متون دیجیتال را موردبررسی قرار داد. وی به این نتیجه رسید که استفاده از این تکنیک در بازیافت موضوعات متون طبقه‌بندی نشده بسیار کاربردی بود. همچنین گاو[2] (۲۰۱۵) در پایان‌نامه دکتری خود به معرفی مدل‌سازی موضوعی بر اساس الگو برای پاک‌سازی و بازیابی اطلاعات پرداخت. نتایج پژوهش وی ارائه روشی نوین برای گسترش سؤال جستجوگر بر اساس الگوهای موضوعی و درنتیجه بهبود بازیابی اطلاعات بود. از دیگر پژوهش‌های صورت گرفته در این راستا مطالعات (Xing Yi & James Allan, 2009; Suominen & Toivanen, 2015; Xu, 2015; Eklund & Nelhans, 2017; Meen & Yongjun, 2018) بوده است که تحلیل‌ها و نتایج حاصل از آن در حوزه تحلیل استنادی و علم‌سنجی نیز استفاده‌شده است.

۳ـ دسته‌بندی مدارک بر اساس موضوعات: آنچه از برونداد کشف موضوعات و برقراری ارتباط بین مدارک از جهت میزان شباهت موضوعی حاصل می‌شود چیزی جز طبقه‌بندی و سازمان‌دهی مدارک و اسناد نخواهد بود؛ که پژوهش‌های (Hu, Fang & Liang, 2014; Lee, Lo, Chen, Lin, & Wang, 2015; Li, He Jun-Qing, Cheng-Long, 2016; Sarioglu, Yadav, & Chio, 2017; Sohrabi, Raeesi vanani,& Baranizade Shineh, 2017; Selvi, 2019) با تمرکز بر این هدف به استفاده از تکنیک مدل‌سازی موضوعی پرداخته‌اند. برای نمونه، هو، فانگ و لیانگ[3] (۲۰۱۴) در مطالعه‌ای تجربی به

1. Cain
2. Gao
3. Hu, Fang, & Liang,

ساخت یک سیستم سازمان‌دهی دانش برای مدارک پروانه ثبت اختراع به کمک مدل‌سازی موضوعی پرداختند. نتایج پژوهش آنها نشان داد که روش مذکور روابط معنایی عمیق بین موضوعات را کشف نموده و به توصیف بهتر زمینه اختراعات درج‌شده در پرونده‌های ثبت اختراع منجر شد. و درنهایت سیستم سازمان‌دهی دانش خلق‌شده به کمک تکنیک مذکور به طبقه‌بندی خودکار اسناد ثبت اختراع و اصلاح دسته‌بندی نتایج جستجو در پایگاه منجر شد. لی، لو، چن، لین و وانگ[1] (۲۰۱۵) نیز در مطالعه‌ای به طبقه‌بندی عناوین خبری به کمک مدل‌سازی موضوعی و الگوریتم LDA پرداختند. نتایج پژوهش آنها نشان داد که روش پیشنهادی نسبت به روش سنتی طبقه‌بندی اطلاعات خبری عملکرد بهتری در بازیابی اطلاعات داشت.

۴ـ کشف الگوهای برجسته و رویدادهای در حال ظهور: با اتکا به گذشته علم، می‌توان آینده علمی یک حوزه تخصصی را پیش‌بینی کرد. بنابراین، داشتن درک و نمایی کلی از چارچوب علمی حوزه موردنظر پژوهشگران را برای رسیدن به هدف‌های پژوهشی یاری می‌کند (سهیلی، شعبانی و خاصه، ۱۳۹۵).

از این‌رو، ارزیابی دقیق موضوعی مقالات می‌تواند شاخصی از روند و گرایش فعالیت‌های پژوهشی هر حوزه را در گذشته جهت آینده‌نگری برای پژوهش‌های بعدی نشان دهد. مدل‌سازی موضوعی به‌عنوان یکی از مطرح‌ترین فن‌های تحلیل متن، ابزاری عالی برای شناخت الگو در انباره‌های بزرگ اطلاعات شناخته‌شده است. بیترمن و فیشر[2] (۲۰۱۸) در مطالعه‌ای به کشف نقاط قوت موضوعی در نشریات حوزه روان‌شناختی پرداختند و برای رسیدن به این هدف از ابزار مدل‌سازی موضوعی استفاده نمودند. آنها از الگوریتم موضوعی LDA برای کشف و استخراج موضوعات نهان در اسناد استفاده کردند و به این نتیجه رسیدند که با استفاده از مدل‌سازی موضوعی، می‌توان موضوعات خاصی را کشف کرد که با سیستم‌های طبقه‌بندی موجود امکان شناسایی آن آسان نیست. فانگ، یانگ، گااو و لی[3] (۲۰۱۸) نیز در مطالعه‌ای به کشف موضوع پژوهش‌ها در میان منابع الکترونیکی کتابخانه پرداختند. هدف آنها کشف موضوعات به‌روز و رایج در پژوهش‌ها بود. آنها به این نتیجه رسیدند که موضوعات کشف‌شده به کمک این روش به‌طور قابل‌توجهی با موضوعات شناسایی‌شده توسط متخصصان سازگار و همخوان است که نشان‌دهنده دقت و ارزش روش مورداستفاده است.

۵ـ خوشه‌بندی مفاهیم حوزه‌های علمی: با انجام فرآیندهای خوشه‌بندی اطلاعات، حیطه

1. Lee, Lo, Chen, Lin, & Wang
2. Bitterman & Fischer
3. Fang, Yang, Gao, & Li

گسترده‌ای از داده‌های پراکنده در گروه‌های مدون و سازمان‌یافته قرار می‌گیرند. برای دسترسی به منابع اطلاعاتی، شناسایی ابعاد و جنبه‌های خاص موضوعی مفاهیم، که با اصطلاح‌های تخصصی مشخص می‌شوند، ضروری است. خوشه‌بندی سبب شده تا گروهی از مفاهیم مشابه در زیر یک رده، با موضوع کلی سازمان‌دهی شوند؛ این فعالیت در دستیابی به اطلاعات مرتبط با موضوع خواسته‌شده تأثیر بسزایی دارد و باعث دسترسی مطلوب به اطلاعات هم موضوع می‌شود (Anthony & Marie, 2007). یکی از بروندادهای اصلی اجرای مدل‌سازی موضوعی خوشه‌بندی مفاهیم مجموعه متون هست. در بررسی مفاهیم حوزه بیوانفورماتیک، لئو و همکارانش[۱] (۲۰۱۶) با بهره‌گیری از تکنیک مدل‌سازی موضوعی به خوشه‌بندی مفاهیم این حوزه پرداختند. یافته‌های آنها حاکی از کارآمدی تکنیک مدل‌سازی موضوعی در خوشه‌بندی مفاهیم حوزه موردنظر بود که باعث ارتقا توان پژوهش‌ها در تفسیر اطلاعات بیولوژیکی نیز می‌شد. سایر مقالات مشابه باهدف خوشه‌بندی مفاهیم در حوزه‌های علمی به کمک الگوریتم‌های مدل‌سازی موضوعی می‌توان به آثار (Castellani Murino, Bellani, Rambaldelli et al, 2010; Masseroli, Chicco, & Pinoli, 2012; Lee, Liu, Kelly,& Tong, 2014; Bisgin, Chen, Wang, Kelly, Hong et al, 2015) اشاره کرد.

۶ـ تحلیل سیر تحول مفهومی در طول دوره‌های زمانی: مفاهیم نه‌فقط به‌دلیل تحول درونی تجربه علمی دچار دگرگونی می‌شوند؛ بلکه تغییر در عقاید و اعمال صاحب‌نظران و پژوهشگران نیز ممکن است موجب دگرگونی مفهومی یک واژه شود. به‌عبارت‌دیگر، کاربرد یک مفهوم در دوره‌های تاریخی به شرایط زمان و مکان آن دوره بستگی دارد. ازاین‌رو، یکی از پژوهش‌های رایج در حوزه‌های علمی مختلف بررسی سیر تحول مفهومی آن حوزه در طول دوره‌های زمانی است. ازآنجاکه در فن‌های متن‌کاوی و ازجمله مدل‌سازی موضوعی استخراج مفاهیم کلیدی مجموعه متون صورت می‌گیرد می‌توان سیر تحول این مفاهیم را با توجه به هم رخدادی آنها در محتوای علوم ملاحظه کرد. برای نمونه، کوراتا[2] (۲۰۱۸) در مطالعه‌ای به تحلیل محتوای مقالات حوزه علم اطلاعات و کتابداری به کمک روش مدل‌سازی موضوعی پرداختند. نتایج پژوهش آنها نشان داد که روش مدل‌سازی موضوعی قادر به شناسایی مفاهیم معمول و تغییرات در روند پژوهش‌های است.

۷ـ تعیین روابط سلسه‌مراتبی مفاهیم یک حوزه یا زمینه خاص علمی: ازنظر سرگل[۳] (۱۹۷۴) رابطه

1. Liu, Hu, & Wang
2. Kurata
3. Soergel

سلسله‌مراتبی برای نشان دادن کل دانش بشری پدید نمی‌آید، بلکه برای نشان دادن سلسله‌مراتب موضوعی در یک زمینه خاص موضوعی به‌منظور بهبود نمایه‌سازی و جستجو پدید آمده است. تعریف تجربی روابط سلسله‌مراتبی را به جزییات «معنی» رابطه مربوط می‌داند. ازنظر وی برای فهم بهتر ماهیت روابط سلسله‌مراتبی که به‌نوبه خود برای توسعه ساختار سلسله‌مراتبی مفید است، بهتر است معنای رابطه با جزییات بیشتر موردتوجه قرار گیرد. هدف از این ساختار، اجازه به کاربر در انتخاب موضوع در عام‌ترین سطح برای افزایش تعداد یافته‌ها یا تقلیل به سطح خاص‌تر برای مهیاکردن دقت بیشتر در نتایج جستجو است (ANSI/ NISO, 2005). پژوهش‌های زیادی با استناد به برونداد تکنیک مدل‌سازی موضوعی که خوشه‌بندی مفاهیم اصلی آثار علمی و رعایت سلسله‌مراتب مفهومی هر خوشه است به تعیین روابط سلسله‌مراتبی مفاهیم آن حوزه پرداخته‌اند (Chen, Anoop, Ashraf, & Deepak, 2016; Chien, 2016; Zhu, Lu, & Liu, 2018).

۸ـ غنی‌سازی فهرست واژگان: فهرست واژگان همچون اصطلاح‌نامه‌ها در محیط‌های اطلاعاتی سنتی و هستی‌شناسی‌ها در فضای اطلاعاتی وب معنایی، یکی از مهم‌ترین ابزارهای سازمان‌دهی منابع اطلاعاتی هستند که با گردآوری اصطلاحات و روابط میان آنها برای نمایش اطلاعات در یک حوزه موضوعی به کار می‌رود که علاوه بر این کارکرد مهم، به‌عنوان ابزاری ارزشمند در خدمت بازیابی اطلاعات است (Aitchison & Shen, 1980). حفظ پویایی و روزآمدی این ابزارها متناسب با ظهور مفاهیم و موضوعات و روابط معنایی جدید حوزه‌های علمی مختلف تأثیر به سزایی در بازنمون مؤثر اطلاعات دارد. از این‌رو، مطالعات مختلفی به بررسی ایجاد تحول در ساختار مفاهیم و روابط به‌موازات ظهور موضوعات و دانش جدید کرده است. نیومن، هاگدوم، چمودوگانتا و اسمیت[1] (۲۰۰۷) به بررسی کارآمدی مدل‌سازی موضوعی در غنی‌سازی فراداده‌های موضوعی پرداختند. هدف آنها ارائه چگونگی عملکرد مدل‌سازی موضوعی در غنی‌سازی فراداده‌های موضوعی و ارتقا کیفیت جستجو در پورتال کتابخانه‌های دیجیتال فدرال امریکا بود. نتایج آنها نشان داد که موضوعات اضافه‌شده به فراداده‌های موضوعی و حاصل از مدل‌سازی موضوعی تأثیر به سزایی در غنی‌سازی مجموعه داشت. رانی، دهار و ویاس[2] (۲۰۱۷) نیز در مطالعه خود به بررسی امکان استفاده از روش مدل‌سازی موضوعی و مقایسه دو الگوریتم مدل‌سازی موضوعی یعنی LSA و LDA در طراحی هستی‌شناسی اصطلاحات تخصصی

1. Newman, Hagedom, Chemudugunta, & Smyth
2. Rani, Dhar, & Vyas

نیمه‌خودکار و طبقه‌بندی موضوعات پرداختند. هدف آنها تعیین ارتباط آماری بین مدارک و اصطلاحات برای ساخت هستی‌شناسی موضوعی و ایجاد گراف هستی‌شناسی با حداقل دخالت انسان و بازیابی معنایی برای موضوع‌ها و تشخیص کلمات در یک حوزه موضوعی بود. نتایج آنها نشان داد که الگوریتم مدل‌سازی موضوعی LDA به‌طور مؤثرتری عمل می‌کند و به ساخت دانش موضوعی و بازیابی معنایی کمک می‌کند.

افزون بر آن، همان‌گونه که اشاره شد، با گسترش پژوهش‌ها و بررسی امکان استفاده از مدل‌سازی موضوعی در حل سایر مسائل، دسته‌بندی پژوهش‌ها را فراتر از هدف‌های یادشده امکان‌پذیر نموده است. برای نمونه، مسعودی و راحتی‌قوچانی (۱۳۹۴) در مطالعه خود برای رفع ابهام از واژگان مبهم فارسی و دارای معانی متعدد، با استخراج ویژگی‌های جدید، مدل موضوعی LDA را پیشنهاد دادند. محمدیان (۱۳۹۳) نیز در پایان‌نامه ارشد خود به ارائه روشی برای شناسایی سرقت‌های علمی در اسناد فارسی بر اساس مدل‌سازی موضوعی پرداخت. وی ضمن دسته‌بندی و بیان ویژگی‌های متدهای تشخیص تقلب علمی بر اساس شباهت‌های متن منبع با متون موجود در پایگاه داده مرجع (تشخیص تقلب خارجی) و استفاده از ویژگی‌های ساختاری سند مشکوک برای کشف تقلب احتمالی (تشخیص تقلب طبیعی) به این نتیجه رسید که روش‌های موجود مشکل هم‌معنی‌ها و چند معنی‌ها را حل نمی‌کند و نیاز به استفاده از معانی مخفی موجود در متن هست. وی بر این اعتقاد بود که استفاده از خوشه‌بندی و پیش‌پردازش‌های صورت گرفته بر روی متن باعث کاهش چشم‌گیر زمان و فضای مورداستفاده برای تشخیص موارد تقلب صورت گرفته بر روی متن می‌شود و این امر امکان تشخیص تقلب در دادگان بزرگ‌تر را امکان‌پذیر می‌نماید.

نتیجه

الگوریتم‌های مدل‌سازی موضوعی روش‌های آماری هستند که کلمات داخل یک متن را تحلیل کرده و از این طریق موضوعات داخل متون را استخراج می‌کنند. همچنین ارتباط این موضوعات با یکدیگر و نیز تغییر آنها در طول زمان را مشخص می‌کنند. مدل‌سازی موضوعی می‌تواند ساختار مفید و قابل تفسیری از متن بدون نیاز به درک روشن و صریح از زبان متن توسط کامپیوتر استخراج کند.

مرور نوشتارها نشان می‌دهد که مدل‌سازی موضوعی با کاهش بعد مدارک یا به‌عبارت‌دیگر خلاصه‌سازی آن در خوشه‌های مفهومی اصلی و بیان نسبت پرداختن به موضوعات هر مدرک تشخیص دقیق‌تری نسبت به حوزه موضوعی منابع ارائه می‌دهد و در کنار سایر فن‌های متن‌کاوی همچون دسته‌بندی و قوانین انجمنی، سنجه مهمی در تجزیه‌وتحلیل محتوای منابع است که در مطالعه شبکه مفهومی یا واژگانی

یک حوزه، رویدادهای موضوعی در حال ظهور و مطالعه سیر تحول مفهومی در طول دوره‌های زمانی به کار گرفته می‌شود. همچنین به‌عنوان فنی در تحلیل محتوا در کشف روابط پنهان میان مفاهیم و روابط معنایی و یکی از روش‌های ایجاد روابط سلسه‌مراتبی مفاهیم در هستی‌شناسی حوزه‌های علمی و زمینه‌های دانش تخصصی کاربرد دارد.

بعلاوه با توجه به ماهیت عملکردی مدل‌سازی موضوعی می‌توان به کمک این تکنیک نشان داد توجهات جامعه علمی، بیشتر به چه موضوعات، زمینه‌ها، مسائل، پژوهش‌ها و نظرات پرداخته است؛ حوزه‌ها و زیر حوزه‌های مختلف علمی کدام‌اند، حوزه‌ها و موضوعات مختلف چه سیر تحولی را پشت سر گذاشته‌اند و احتمالاً در آینده چه زمینه‌های موضوعی موردتوجه قرار خواهند گرفت و همچنین چه رابطه مفهومی بین حوزه‌ها و زیر حوزه‌های مفهومی وجود دارد که برونداد و تأثیر نهایی آن بر سازمان‌دهی و ذخیره و بازیابی بهینه اطلاعات خواهد بود. همچنین می‌توان ادعا کرد که مدل‌سازی موضوعی در راستای خودکارسازی استخراج موضوع و کشف موضوعات نهان موجود در منبع می‌تواند در راستای تقویت و روزآمدسازی نظام‌های نوین سازمان‌دهی منابع اطلاعاتی به کار رود و تمامی این موارد به‌عنوان فضاهای خالی پژوهش‌ها و پیشنهاد برای پژوهش‌های آتی قابل بررسی است.

منابع

سهیلی، فرامرز؛ شعبانی، علی؛ خاصه، علی‌اکبر (۱۳۹۵). ساختار فکری دانش در حوزه رفتار اطلاعاتی: مطالعه هم واژگانی. *تعامل انسان و اطلاعات*، ۲(۴)، ۲۱-۳۶. بازیابی شده در ۲۳ تیر ۱۴۰۰ از https://hii.khu.ac.ir/article-1-2446-fa.pdf

محمدیان، بنفشه (۱۳۹۳). *شناسایی سرقت علمی در اسناد فارسی بر اساس مدل‌سازی موضوعی*. پایان‌نامه کارشناسی ارشد. دانشگاه خوارزمی. استاد راهنما میر محسن پدرام. بازیابی شده در ۲۳ تیر ۱۴۰۰ از https://ganj-old.irandoc.ac.ir/articles/800558

مسعودی، بابک؛ راحتی‌قوچانی، سعید (۱۳۹۴). رفع ابهام معنایی واژگان مبهم فارسی با مدل موضوعی LDA. *پردازش علائم و داده‌ها*. ۴(۲۶)، ۱۱۷-۱۲۵. بازیابی شده در ۲۳ تیر ۱۴۰۰ از: http://jsdp.rcisp.ac.ir/article-1-58-fa.html

Arora, M., Kanjilal, U., Varshney, D(2010). Using social network analysis for information retrieval for mapping information and knowledge flows. *Consultation Workshop on Knowledge Management, IMT Ghaziabad*. Retrieved in 10 july 2020 from SSRN: https://ssrn.com/abstract=2162039

Aitchison, J. and Shen, S. (1980). Logistic normal distributions: Some properties and uses. *Biometrika* (67) 261–272. Retrieved in 10 jun 2020 from https://www.jstor.org/stable/2335470

Alghamdi, R., Alfalqi, Kh. (2015). A survey of topic modeling in text mining. I. J.

ACSA,(1), 147-153. Retrieved jun30, 2020 from http://Doi.org/ 10.14569/IJACSA.2015.060121

American National Standards Institute/National Information Standards Organization (ANSI/NISO). (2005). Guidelines for the construction, format, and management of monolingual thesauri. *Bethesda: NISO Press*. Retrieved in 10 september 2020 from http://www.niso.org/standards/

Anoop, V.s., Ashraf, S., Deepak, P. (2016). Unsupervised concept hierarchy learning: topic modeling guided approach. *Procedia computer science*. (89): 386- 394. Retrieved in 11 september from: https://www.sciencedirect.com/science/article/pii/S1877050916311516

Anthony, A.; DesJardins, M. (2007). "Data clustering with a relational push- pull mod-el". *Association for the Advancement of ArtificialIntelligence*. Retrieved Nov.30, 2019 from: http://maple.cs.umbc.edu/papers/anthonya-Cluster-ingRPPM.pdf

Baeza-Yates, B., Ribeiro-Neto. *Modern Information Retrieval*. ACM Press, New York, 1999.

Bisgin H., Chen M., Wang Y., Kelly R., Hong F. et al. (2013) A systems approach for analysis of high content screening assay data with topic modeling. *BMC Bioinform* 14(Suppl 14):1–10 .Retrieved Nov.30, 2019 from https://www.researchgate.net/publication/258169928_A_Systems_Approach_for_Analysis_of_High_Content_Screening_Assay_Data_with_Topic_Modeling

Bitterman, A.; Fischer, A. (2018). How to identify hot topics in psychology using topic modeling. *Zeitschrift fur psychologie*. 226(1) 3-13 .Retrieved Nov.30, 2020 from https://www.researchgate.net/publication/321171283_How_to_Identify_Hot_Topics_in_Psychology_Using_Topic_Modeling

Blei, D., Jordan, M. (2003), "Latent dirichlet allocation.*" Journal of Machine Learning Research*, (3), 993–1022. Retrieved jun30, 2020 from https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Blei, D. & Lafferty, J. (2007). A correlated topic model of science. *The annual of applied statistics*,1(1), 17-35. Retrieved jun30, 2020 from http://Doi.org/ 10.1214/07-AOAS114

Boyd-Graber, J., Blei, D.M., Zhu, X. (2007). A topic model for word sense disambiguation. In: *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. 1024–1033. Retrieved jun30, 2020 from https://www.researchgate.net/publication/221013017_A_Topic_Model_for_Word_Sense_Disambiguation

Castellani U., Perina A., Murino V., Bellani M., Rambaldelli G. et al. (2010) Brain morphometry by probabilistic latent semantic analysis. *Int Conf Med Image Comput Computer Assist Intervent* (13)177–184: Retrieved jun30, 2020 from http://Doi.org/ 10.1007/978-3-642-15745-5_22

Chen, J; Zhu, J; Lu, J; Liu, S. (2018). Scalable training of hierarchical topic modelis. *Proceeding of VLDB endowment*, 11(7), 826- 839. Retrieved jun30, 2020 from: https://www.researchgate.net/publication/325304116_Scalable_training_of_hierarchical_topic_models.

Chien, Jt. (2016). Hierarchical theme and topic modeling. *IEEE trans neural netw learn syst*.27(3): 565-578.Retrieved jun20, 2020 from:

https://www.researchgate.net/publication/274394886_Hierarchical_Theme_and_Topic_Modeling.

Ding, Y., Chowdhhury, G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using coword analysis. *Information Processing And Management*, 37, 817-842. Retrieved nov30, 2020 from: https://www.sciencedirect.com/science/article/abs/pii/S0306457300000510

Eklund, J.; Nehans, G. (2017). Topic modeling approaches to aggregated citiation data. *Science, technology and innovation indicators*. Retrieved in 10 april from: https://pdfs.semanticscholar.org/ce33/50eb3de81f84b07177bd856e326c9bc4c7f5.pdf.

Fang, D., Yang, H., Gao, B. and Li, X. (2018), "Discovering research topics from library electronic references using latent Dirichlet allocation", *Library Hi Tech*, 36(3), 400-410. Retrieved in 20 april 2020 from:https://doi.org/10.1108/LHT-06-2017-0132.

Fayyad, U.; Piatestky-Shapiro, G. (2010). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37. Retrieved in 10 april 2020 from https://ojs.aaai.org//index.php/aimagazine/article/view/1230

Figuerola, C.G., García Marco, F.J. & Pinto, M. (2017) *Scientometrics* 112: 1507. Retrieved 10 Feb 2020 from https://doi.org/10.1007/s11192-017-2432-9.

Gao, Y. (2015). Pattern-based topic modeling and its application for information filtering and information retrieval. PhD thesis. Computer Department. University of technology, Queensland. Retrieved 11 April 2020 from: https://eprints.qut.edu.au/83982/1/Yang_Gao_Thesis.pdf

Gertisio, Ch. & Dussauchoy, A. (2010). “Knowledge Discovery from Industrial Data base”. *Journal of Intelligent Manufacturing*, 15, 29-37. Retrieved 20 Feb 2020 from https://www.researchgate.net/publication/227069567_Knowledge_discovery_from_industrial_databases

Hagen, L. (2018). Content analysis of e-petition with topic modeling: how to train and evaluate LDA models? *Information processing & management*,54(6), 1292-1307. Retrieved 10 April 2019 from https://www.sciencedirect.com/science/article/abs/pii/S0306457317307240

Hofmann, T. (1999) "Probabilistic latent semantic analysis," *Proceedings of the 15th Conference on Uncertainty in AI*. Retrieved 2 Feb 2020 from http://canterbury.eblib.com.au/patron/fullrecord.aspx?p=1687540

Hu, Z., Fang, S. & Liang, T. (2014)Emprical study of constructing a knowledge organization system of patent documents using topic modeling, *Scientometrics*, 100(3) 787- 799, Retrieved 11 Nov 2020 from: https://link.springer.com/article/10.1007%2Fs11192-014-1328-1

Hwang, S.Y., Wei, C.P., Lee, C.H., & Chen, Y.S. (2017). Coauthor ship network based literature recommendation with topic model. Online Information Review, 41(3), 318-336. Retrieved 11 Nov 2020 from https://www.emerald.com/insight/content/doi/10.1108/OIR-06-2016-0166/full/html

Cain, J. O. (2016): Using Topic Modeling to Enhance Access to Library Digital Collections, *Journal of Web Librarianship*. Retrieved 11 Nov 2020 from DOI: 10.1080/19322909.2016.1193455

Kontostathis, A., and Pottenger, W., (2006) “A framework for understanding Latent Semantic Indexing (LSI) performance”, *Information Processing and Management*, 42 (1), 56- 73. Retrieved 9 Nov 2020 from

https://www.sciencedirect.com/science/article/abs/pii/S0306457304001529

Kostoff, R. N., Eberhart, H.J, & Toothman. D. R. (1997). Database Tomography for information retrieval. *Journal of Informution Science* 23 (4): 301-311. Retrieved 9 March 2020 from https://www.sciencedirect.com/science/article/abs/pii/S0306457304001529

Kurata, K. (2018). Analyzing library and information science full-text articles using a topic modeling approach. *81Annual meeting of the association for information science & technology I nVancouvar of Canada* (10-14, November, 2018). Retrieved 14 March 2020 from: https://www.researchgate.net/publication/330812928_Analyzing_library_and_information_science_full-text_articles_using_a_topic_modeling_approach

Lee, M., Liu, Z., Kelly, R., Tong, W. (2014) Of text and gene—using text mining methods to uncover hidden knowledge in toxicogenomics. *BMC Syst Biol* 8(1):1–11. Retrieved 14 March 2020 from https://www.researchgate.net/publication/264797394_Of_text_and_gene_-_using_text_mining_methods_to_uncover_hidden_knowledge_in_toxicogenomics

Lee, Y. S., Lo, R., Chen, C. Y., Lin, P. C., & Wang, J. C. (2015). News topics categorization using latent dirichlet allocation and sparse representation classifier. *In 2015 IEEE international conference consumer Electronics-Taiwan (ICCE-TW)* (pp. 136–137). Taiwan. Retrieved 14 March 2020 from https://ieeexplore.ieee.org/document/7216819

Li, P; He Jun-Qing; Ma, Ch(2016). Short text classification based on Latent Topic Modeling and word embedding. *Joint international conference on artificial intelligence and computer engineering*. Retrieved 14 March 2020 from https://www.researchgate.net/publication/313541675_Short_text_classification_based_on_LDA_topic_model

Liu, G., Hu, J., & Wang, H. (2012). A Co-word analysis of digital library field in china. *Scientometrics* 91(1), 203-217. Retrieved 1 March 2020 from https://link.springer.com/article/10.1007/s11192-011-0586-4

Masseroli, M., Chicco, D., Pinoli, P. (2012) Probabilistic latent semantic analysis for prediction of gene ontology annotations. *The 2012 international joint conference on neural networks (IJCNN)*, 1–8. Retrieved 1 April 2020 from https://www.researchgate.net/publication/261087004_Probabilistic_Latent_Semantic_Analysis_for_prediction_of_Gene_Ontology_annotations

Meen, Ch. & Yongjun, Zh. (July 18th 2018). Scientometrics of Scientometrics: Mapping Historical Footprint and Emerging Technologies in Scientometrics, *Scientometrics*,. Retrieved 12 April 2020 from: https://www.intechopen.com/books/scientometrics/scientometrics-of-scientometrics-mapping-historical-footprint-and-emerging-technologies-in-scientome

Nadezhda, Y & Aleksey, F (2018). Improving the quality of information retrieval using syntactic analysis of search query. Retrieved 23 April 2020 from:https://www.semanticscholar.org/paper/Improving-the-Quality-of-Information-Retrieval-of-Yarushkina-ilippov/d0955103ee4e4cd78a0d24f880a1cda7f3b35d5e

Neff, M. W., & Corley, E. A. (2009). 35 years and 160,000 articles: a bibliometric exploration of the evolution of ecology. *Scientometrics*, 80(3), 657–682. Retrieved 23 April 2020 from

https://www.researchgate.net/publication/257662787_35_years_and_160000_articles_A_bibliometric_exploration_of_the_evolution_of_ecology

Papadimitriou, C, H., Raghavan, P., Tamaki, H., and Vempala, S(1998). Latent semantic indexing: A probabilistic analysis. *Journal of Computer and system sciences*. 61(2), 217-235. Retrieved 23 Nov 2020 from https://www.sciencedirect.com/science/article/pii/S0022000000917112

Newman, D., Hagedorn, K., Chemudugunta, C., & Smyth, P. (2007). Subject metadata enrichment using statistical topic models. *JCDL*. Retrieved 23 Nov 2020 from file:///C:/Users/Ariyana/AppData/Local/Temp/Subject_Metadata_Enrichment_using_Statistical_Topi.pdf

Rani, M; Dhar, A; Vyas, O.P(2017). Semi- Automatic terminology ontology learning based on topic modeling. *Engineering Application of Artificial Intelligence*, 63, 108-125. Retrieved 15 Nov 2020 from https://www.sciencedirect.com/science/article/abs/pii/S0952197617300891

Rokaya, M. (2008). Ranking of field association terms using co-word analysis. Information Processing & Management, 44(2) 738-755. Retrieved 26 Nov 2020 from https://www.sciencedirect.com/science/article/abs/pii/S0306457307001227

Sanandres, E; Madariaga, C; Abello, R(2018). Topic modelling of twitter conversations. Retrieved 26 Nov 2020 from: https://www.researchgate.net/publication/326450126_Topic_Modeling_of_Twitter_Conversations/citations

Sarioglu, E.; Yadav, K.; Chio, H. (2017). Topic modeling based classification of clinical report. *Association for computational linguistics*, 67-73. Retrieved 22 Nov 2020 from: http://aclweb.org/anthology/P13-3010.

Selvi, M. (2019). Classification of Medical Dataset Along with Topic Modeling Using LDA. In: Nath V., Mandal J. (eds) Nanoelectronics, Circuits and Communication Systems. Lecture Notes in Electrical Engineering, 511. *Springer*. Retrieved 22 Nov 2020 from https://doi.org/10.1007/978-981-13-0776-8_1

Soergel, D(1974) .*Indexing language and thesauri: construction and maintenance*. Los Angeles, CA: Melville.

Sohrabi, B.; Raeesi vanani, I.; Baranizade Shineh, M. (2017). Topic Modeling and classification of cyberspace papers using text mining. *Cyberspace studies*, 2(1), 103-125. Retrieved 1 Nov 2020 from http://ensani.ir/fa/article/380869/topic-modeling-and-classification-of-cyberspace-papers-using-text-mining

Sparck Jones, K. (1971). *Automatic Keyword Classification for Information Retrieval*. Butterworths: London.

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. *Handbook of latent semantic analysis*, 427(7), 424-440. Retrieved 1 April 2020 from https://psycnet.apa.org/record/2007-04818-021

Suominen, A.; Toivanen, H. (2015). Map of Science with topic modeling: Comparison of unsupervised learning and human- assigned subject classification. *Journal of association for information and technology*,67(10):2464–2476. Retrieved 12 April 2020 from https://dl.acm.org/doi/10.1002/asi.23596

Vickery, B.; Vickery, A. (1987). *Information science in theory and practice*. Berlin, New York: K. G. Saur Retrived 10 Jun from: https://doi.org/10.1515/9783598440083

Wittgenstein, L. (2010). *Philosophical investigations*. John Wiley & Sons.

Xing, Y., & James A (2009). *A Comparative Study of Utilizing Topic Models for Information Retrieval. In Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval (ECIR '09).* Springer-Verlag, Berlin, Heidelberg, 29–41. Retrived 10 Jun from DOI:https://doi.org/10.1007/978-3-642-00958-7_6

Xu, Z. (2015) Doctoral dissertations in Chinese Interpreting Studies: A scientometric survey using topic modeling. *FORUM. Revue internationale d'interprétation et de traduction / International Journal of Interpretation and Translation.*13(1),131-165. Retrived 12 Jun from https://doi.org/10.1075/forum.13.1.07xuz

Topic Modeling and its Application in Research: A Review of Specialized Literature

Fateme Zarmehr
Ali Mansouri*
Hosein Karshenas

Isfahan University

Abstract

Introduction: Topic modeling is one of the text mining techniques that allows you to discover unknown topics in a collection of documents, interpret documents based on these topics, and use these interpretations to organize, summarize, and search for texts automatically. Familiarity with the concept and technique of topic modeling, and its application in discovering topics and organizing information is one of the main goals of this research.

Methodology: The present study is a review-analytical type in which, while introducing topic modeling, it has categorized and reviewed the applications of this technique based on its performance and provided a sample of research that has used this technique.

Findings: Topic modeling algorithms is used not only in addition to the three main objectives of discovering hidden topics, interpreting documents based on topics, and finally organizing and classifying texts, but also is used in discovering hidden topics and relationships in the fields of science, information retrieval, categorizing documents based on topics, discovering outstanding patterns and emerging events, clustering the concepts of scientific fields, analyzing the course of conceptual evolution during historical periods, determining the hierarchical relationships of concepts. A specific scientific field or field and vocabulary enrichment.

Conclusion: Topic modeling based on machine learning and artificial intelligence knowledge has been proposed as one of the new approaches to organizing information resources and serious studies are being conducted in this field. Therefore, by using topic modeling algorithms in order to automate the extraction of the subject and discover the hidden issues in the source, it is possible to strengthen and update the new systems of organizing information resources.

Keywords: Text mining, Topic Modeling, Subject Discovery, Information Organization, Subject Diagnosis, Subject Allocation

*. a . mansouri@edu.ui.ac.ir